Sequential Detection and Isolation of a Correlated Pair

Anamitra Chaudhuri
Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, Illinois 61820
E-mail: ac34@illinois.edu

Georgios Fellouris
Department of Statistics
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Champaign, Illinois 61820
E-mail: fellouri@illinois.edu

***Abstract*—The problem of detecting and isolating a correlated pair among multiple Gaussian information sources is considered. It is assumed that there is at most one pair of correlated sources and that observations from all sources are acquired sequentially. The goal is to stop sampling as quickly as possible, declare upon stopping whether there is a correlated pair or not, and if yes, to identify it. Specifically, it is required to control explicitly the probabilities of three kinds of error: false alarm, missed detection, wrong identification. We propose a procedure that not only controls these error metrics, but also achieves the smallest possible average sample size, to a first-order approximation, as the target error rates go to 0. Finally, a simulation study is presented in which the proposed rule is compared with an alternative sequential testing procedure that controls the same error metrics.**

I. Introduction

The quick detection and isolation of the correlation structure in a p-variate Gaussian random vector is of interest in many areas of science and engineering, such as environmental monitoring [1], blind source separation [2], biometric authentication [3], sensor networks [4], fault detection in power grid [5], neural coding [6]. When $p = 2$, we have the problem of sequential testing regarding the correlation coefficient of a bivariate Gaussian distribution. This problem has been considered in [7], [8], [9], [10], where the proposed tests reduce to certain Sequential Probability Ratio Tests [11], and also in [12], where a repeated generalized likelihood ratio test with a fixed maximum sample size was proposed. In the general multivariate case $(p > 2)$, it is natural to consider a sequential multiple testing problem in the spirit of [13], [14], [15], [16], [17]. Such a formulation was proposed in [18] under the assumption that only one component/source can be observed at each sampling instance and that observations from dependent components exhibit the same correlation even when they are taken far apart in time.

In this paper we assume that we observe at each sampling instance all components of a Gaussian random vector and that *at most* one pair of components is correlated. The goal is to stop sampling as quickly as possible, declare upon stopping whether there is a correlated pair or not, and if yes, to identify it. Specifically, we need to control explicitly the probabilities of three kinds of error: detecting a correlated pair when there is none (*false alarm*), failing to detect any correlated pair when there is one (*missed detection*), claiming correctly that there is a correlated pair but identifying it incorrectly (*wrong isolation*). The main contribution of this work is that we propose a procedure that controls explicitly these error metrics, and most importantly it achieves the smallest possible average sample size, to a first-order approximation, as the target error rates go to 0.

The proposed rule is inspired by the "gap-intersection" rule in [16], however there are some important differences between this work and the current paper. First, the test statistics that correspond to the individual hypotheses in [16] are independent, which is not the case in our setup. Second, in [16] there are only two kinds of error under control, whereas here we explicitly control three error metrics. Third, [16] deals with the simultaneous testing of binary *simple* hypotheses, whereas here we are interested in the simultaneous testing of simple nulls versus *two-sided* alternatives.

Finally, we also compare numerically the proposed procedure with a modified version of the "intersection" rule, proposed in [17], which takes into account the prior information that there is at most one correlated pair. This procedure has two free parameters, which nevertheless can be selected so that the three error constraints are satisfied. Our simulation study suggests that the proposed procedure performs significantly better, at least when the constraint on the probability of wrong isolation is the strictest.

The remainder of the paper is organized as follows: In Section II we formulate the problem mathematically. In Section III we introduce the proposed rule and in Section IV we establish its asymptotic optimality. In Section V we present the results of the simulation study.

II. Problem formulation

We consider p information sources, $\{X_i(t) : t \in \mathbb{N}\}$, $i = 1, \dots p$, each generating a sequence of i.i.d. standard Gaussian random variables, i.e., $X_i(t) \sim \mathcal{N}(0, 1)$ for every $t \in \mathbb{N} \equiv \{1, 2 \dots\}$ and $1 \le i \le p$. The correlation between any two sources is assumed to be constant over time, which means that if we denote by $\mathcal{E}$ the set of all (unordered) pairs, i.e., $\mathcal{E} := \{(i, j) : 1 \le i < j \le p\}$, and $e = (k, l) \in \mathcal{E}$, then there is a number $\rho_e \in (-1, 1)$ such that

$$\text{Corr}(X_k(t), X_l(t)) = \rho_e \quad \forall\, t \in \mathbb{N}.$$

We are interested in the case that there is *at most one* strongly correlated pair of sources and the problem is to detect whether such a pair exists and, if this is indeed the case, to identify it. Specifically, given a user-specified value $\rho_* \in (0,1)$ that represents a correlation value large enough to be of interest, we want to test, *for each pair* $e \in \mathcal{E}$, the null hypothesis $\rho_e = 0$ against the *two-sided* alternative $|\rho_e| = \rho_*$, when *at most one* of the $K := |\mathcal{E}| = \binom{p}{2}$ nulls should be rejected.

We assume that the data from all sources become available sequentially and the goal is to stop sampling and make a decision as soon as possible. Thus, the σ-algebra generated by the observations in the first n sampling instances is $\mathcal{F}_n := \sigma(X(1), \ldots, X(n))$, where $X(n) := (X_1(n), \ldots, X_p(n))$. A *sequential* test for this problem is a pair (τ, d) that consists of an $\{\mathcal{F}_n\}$-stopping time, τ, at which we stop sampling from all the sources, and an $\mathcal{F}_\tau$-measurable decision rule $d := (d_e, e \in \mathcal{E})$, each element of which is a Bernoulli random variable, with the understanding that for each $e \in \mathcal{E}$ the sources in e are declared to be correlated (resp. independent) upon stopping when $d_e = 1$ (resp. $d_e = 0$). With an abuse of notation, we also use d to denote the subset of pairs which are declared to be correlated upon stopping. Since we focus on the case that there is at most one correlated pair, we restrict ourselves to decision rules for which at most one element can be equal to 1.

We are interested in controlling below α the probability of detecting a correlated pair when there is none (*false alarm*), below β the probability of failing to detect a correlated pair when there is one (*missed detection*), below γ the probability of incorrectly identifying the correlated pair when there is one (*wrong isolation*), where $\alpha, \beta, \gamma \in (0,1)$ are user-specified thresholds. To be more specific, we denote by P_{e+} (resp. P_{e-}) the underlying probability measure when the pair $e \in \mathcal{E}$ has correlation ρ_* (resp. $-\rho_*$), while all other sources are independent. We also denote by P_0 the underlying probability measure when all sources are independent. We denote by $\Delta(\alpha, \beta, \gamma)$ the class of sequential tests (τ, d) for which $P_0(d \neq \emptyset) \leq \alpha$ and also for every $e \in E$

$$\begin{aligned} & P_{e+}(d = \emptyset),\ P_{e-}(d = \emptyset) \leq \beta, \\ \text{and} \quad & P_{e+}(d \neq \emptyset, d \neq \{e\}), P_{e-}(d \neq \emptyset, d \neq \{e\}) \leq \gamma. \end{aligned}$$

The main result of this paper is that we obtain a sequential test that not only controls these error metrics, but also achieves the smallest possible expected sample size among all sequential tests in $\Delta(\alpha, \beta, \gamma)$, to a first-order asymptotic approximation as $\alpha, \beta, \gamma \to 0$, under P_0 *and* P_{e+}, P_{e-}, *for every* $e \in \mathcal{E}$.

A. Notations and Statistics

For each $e \in \mathcal{E}$ we denote by $\Lambda_{e+}(n)$ (resp. $\Lambda_{e-}(n)$) the likelihood ratio of P_{e+} (resp. P_{e-}) versus P_0 after the first n sampling instances, i.e.,

$$\Lambda_{e+}(n) := \frac{dP_{e+}}{dP_0}(\mathcal{F}_n), \quad \Lambda_{e-}(n) := \frac{dP_{e-}}{dP_0}(\mathcal{F}_n), \qquad (1)$$

which take the following form

$$\begin{aligned} \Lambda_{e+}(n) &= z_n \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \mathbf{X}_e^T(t)(\Sigma_+^{-1} - I_2)\mathbf{X}_e(t) \right\} \\ \Lambda_{e-}(n) &= z_n \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \mathbf{X}_e^T(t)(\Sigma_-^{-1} - I_2)\mathbf{X}_e(t) \right\}, \end{aligned} \qquad (2)$$

where $z_n := (1-\rho_*^2)^{-n/2}$, $\mathbf{X}_e(t)$ denotes the components of $X(t)$ that correspond to the pair e, I_p is the $p \times p$ identity matrix, and

$$\Sigma_+ := \begin{pmatrix} 1 & \rho_* \\ \rho_* & 1 \end{pmatrix}, \quad \Sigma_- := \begin{pmatrix} 1 & -\rho_* \\ -\rho_* & 1 \end{pmatrix}. \qquad (3)$$

Since for each $e \in \mathcal{E}$ we have a two-sided testing problem, we also introduce the mixture probability measure

$$P_e := (P_{e+} + P_{e-})/2, \qquad (4)$$

and define the following mixture likelihood ratio

$$\Lambda_e(n) := \frac{dP_e}{dP_0}(\mathcal{F}_n) = (\Lambda_{e+}(n) + \Lambda_{e-}(n))/2. \qquad (5)$$

We use the following notation for the ordered mixture likelihood ratio statistics at time n:

$$\Lambda^{(1)}(n) \geq \cdots \geq \Lambda^{(K)}(n),$$

and we denote by $i_1(n), \ldots, i_K(n)$ the corresponding pairs, i.e.,

$$\Lambda^{i_k}(n) \equiv \Lambda^{(k)}(n) \quad \text{for every} \quad k \in \{1, \ldots K\}.$$

We also introduce the corresponding log-likelihood ratio processes:

$$\begin{aligned} & Z_{e+}(n) := \log \Lambda_{e+}(n),\ Z_{e-}(n) := \log \Lambda_{e-}(n), \\ & Z_e(n) := \log \Lambda_e(n). \end{aligned} \qquad (6)$$

The mixture log-likelihood ratio, Z_e, is not a random walk under P_{e+}, as it is the case for Z_{e+}, but we have the following decomposition

$$\begin{aligned} & Z_e(n) = Z_{e+}(n) + \xi_e(n) - \log 2, \quad \text{where} \\ & \xi_e(n) := \log\left(1 + \exp\{Z_{e-}(n) - Z_{e+}(n)\}\right). \end{aligned} \qquad (7)$$

We note that since $\xi_e(n) > 0$, $Z_e(n) \geq Z_{e+}(n) - \log 2$. Moreover, for every $e' \in \mathcal{E}$ with $e' \neq e$ and $n \in \mathbb{N}$ we have

$$Z_e(n) - Z_{e'}(n) \geq Z_{e+}(n) - Z_{e'+}(n) - \xi_{e'}(n). \qquad (8)$$

We denote by E_{e+} (resp. E_{e-}) expectation under P_{e+} (resp. P_{e-}) and by E_e (resp. E_0) expectation under P_e (resp. P_0). If P represents any of these measures and E is expectation under P, then for any event Γ we use the following notation:

$$E[Y; \Gamma] := \int_\Gamma Y dP.$$

Finally, for each $e \in \mathcal{E}$ we introduce the following Kullback-Leibler information numbers

$$\begin{aligned} D_0 &:= E_0[-Z_{e+}(1)] = E_0[-Z_{e-}(1)], \\ D_1 &:= E_{e+}[Z_{e+}(1)] = E_{e-}[Z_{e-}(1)]. \end{aligned} \qquad (9)$$

III. Proposed Procedure

In this section we introduce the proposed sequential testing procedure and we show how it can be designed to guarantee the desired error control. We suggest that sampling be stopped as soon as either all mixture likelihood ratio statistics are below $1/A$, or the largest mixture likelihood ratio statistic is above B and *at the same time* larger by a factor C than the second largest mixture likelihood ratio statistic, where $A, B, C > 1$ should be determined from the error constraints. Specifically, the stopping rule of the proposed procedure is

$$\begin{aligned} &\tau_* := \min\{\tau_1, \tau_2\}, \quad \text{where} \\ &\tau_1 := \inf\{n \geq 1 : \Lambda^{(1)}(n) \leq 1/A\}, \\ &\tau_2 := \inf\{n \geq 1 : \Lambda^{(1)}(n) \geq B, \Lambda^{(1)}(n)/\Lambda^{(2)}(n) \geq C\}. \end{aligned}$$

When $\tau_1 < \tau_2$, we declare upon stopping that there is no correlated pair of sources. When $\tau_2 < \tau_1$, we declare that the the correlated pair is the one that corresponds to the largest mixture likelihood ratio statistic, i.e.,

$$d_* := \begin{cases} \emptyset & \text{if } \tau_1 < \tau_2 \\ i_1(\tau_*) & \text{if } \tau_2 < \tau_1. \end{cases} \tag{10}$$

Since $\lim_{n\to\infty} \Lambda_e(n) = 0$ almost surely under P_0 for every $e \in \mathcal{E}$, τ_1 is almost surely finite under P_0. Moreover, for any $e, e' \in \mathcal{E}$ with $e \neq e'$ it is clear that $\lim_{n\to\infty} \Lambda_e(n) = \infty$ and $\lim_{n\to\infty} \Lambda_{e'}(n) = 0$ almost surely under P_{e+} and P_{e-}, therefore τ_2 is almost surely finite under P_{e+} and P_{e-}. We conclude that, for any choice of A, B, C, the proposed testing procedure, (τ_*, d_*), terminates almost surely under P_0 and under P_{e+}, P_{e-} for every $e \in \mathcal{E}$. The following theorem shows how to select A, B, C for (τ_*, d_*) to satisfy the three error constraints. We recall that K represents the number of all possible pairs, i.e., $K = \binom{p}{2}$.

Theorem 3.1: For any $A, B, C > 1$, we have

$$P_0(d_* \neq \emptyset) \leq K/B, \tag{11}$$

$$P_{e+}(d_* = \emptyset) = P_{e-}(d_* = \emptyset) \leq 1/A, \tag{12}$$

$$\begin{aligned} P_{e+}(d_* \neq \emptyset, d_* \neq \{e\}) &= P_{e-}(d_* \neq \emptyset, d_* \neq \{e\}) \\ &\leq (K-1)/C. \end{aligned} \tag{13}$$

In particular, $(\tau_*, d_*) \in \Delta(\alpha, \beta, \gamma)$ when

$$A = \frac{1}{\beta}, \quad B = \frac{K}{\alpha} \text{ and } C = \frac{K-1}{\gamma}. \tag{14}$$

Proof: Fix $A, B, C > 1$. When a pair is declared to be correlated upon stopping, $\Lambda_e(\tau_*) \geq B$ for some $e \in \mathcal{E}$, which means that

$$\{d_* \neq \emptyset\} \subseteq \bigcup_{e\in\mathcal{E}} \Gamma_e, \quad \text{where } \Gamma_e := \{\Lambda_e(\tau_*) \geq B\}.$$

Applying Boole's inequality and Wald's likelihood ratio identity we obtain

$$P_0\big(d_* \neq \emptyset\big) \leq \sum_{e\in\mathcal{E}} P_0\big(\Gamma_e\big) \leq \sum_{e\in\mathcal{E}} E_e\left[\frac{1}{\Lambda_e(\tau_*)}; \Gamma_e\right] \leq \frac{K}{B},$$

which proves (11). The equalities in (12)-(13) follow by the symmetry of the statistics in (2). For the inequality in (12) we observe that there is a missed detection under P_{e+} when the event $\Gamma'_e := \{\Lambda_e(\tau_*) \leq 1/A\}$ occurs, therefore

$$P_{e+}\big(d_* = \emptyset\big) \leq P_{e+}\big(\Gamma'_e\big).$$

By the symmetry of the statistics in (2) it follows that

$$P_{e+}(\Gamma'_e) = P_{e-}(\Gamma'_e) = P_e(\Gamma'_e).$$

From these two relationships and another application of Wald's likelihood ratio identity we obtain

$$P_{e+}\big(d_* = \emptyset\big) \leq P_e\big(\Gamma'_e\big) = E_0[\Lambda_e(\tau_*); \Gamma'_e] \leq 1/A,$$

which completes the proof of (12).

Finally, when there is a wrong identification under P_{e+}, the event $\Gamma_{e',e} := \{\Lambda_{e'}(\tau_*)/\Lambda_e(\tau_*) \geq C\}$ occurs for some $e' \in \mathcal{E}$ such that $e' \neq e$. Therefore,

$$\big\{d_* \neq \emptyset, d_* \neq \{e\}\big\} \subseteq \bigcup_{e'\neq e} \Gamma_{e',e},$$

and by Boole's inequality we have

$$P_{e+}\big(d_* \neq \emptyset, d_* \neq \{e\}\big) \leq \sum_{e'\neq e} P_{e+}\big(\Gamma_{e',e}\big).$$

By the symmetry of the statistics in (2) it follows that

$$P_{e+}(\Gamma_{e',e}) = P_{e-}(\Gamma_{e',e}) = P_e(\Gamma_{e',e}),$$

and applying Wald's likelihood ratio identity again we obtain

$$P_e(\Gamma_{e',e}) = E_{e'}\left[\frac{\Lambda_e(\tau_*)}{\Lambda_{e'}(\tau_*)}; \Gamma_{e',e}\right] \leq 1/C.$$

Combining the last three relationships we obtain the inequality in (13). ∎

IV. Asymptotic Optimality

In this section we establish a *non-asymptotic* lower bound on the expected sample size of an arbitrary procedure in $\Delta(\alpha, \beta, \gamma)$ under P_0 and P_{e+}, P_{e-}, $e \in \mathcal{E}$, and then we show that all these lower bounds are attained by the proposed rule to a first-order approximation as $\alpha, \beta, \gamma \to 0$. To state the lower bounds we need to define the following function:

$$h(x, y) := x \log\left(\frac{x}{1-y}\right) + (1-x) \log\left(\frac{1-x}{y}\right) \tag{15}$$

where $x, y \in (0, 1)$. Moreover, we set $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$.

Lemma 4.1: If $\alpha, \beta, \gamma \in (0, 1)$ such that $\alpha + \beta < 1$ and $\beta + 2\gamma < 1$, $e \in \mathcal{E}$, and $(\tau, d) \in \Delta(\alpha, \beta, \gamma)$, then

$$\begin{aligned} E_0[\tau] &\geq \frac{h(\alpha, \beta)}{D_0} \\ E_{e+}[\tau], E_{e-}[\tau] &\geq \frac{h(\beta, \alpha)}{D_1} \bigvee \frac{h(\beta + \gamma, \gamma) \vee h(\gamma, \beta + \gamma)}{D_0 + D_1}. \end{aligned}$$

Proof: We start with the proof of the first lower bound. Without loss of generality, we assume that $E_0[\tau] < \infty$. Under

P_0, the log-likelihood ratio process $\{-Z_{e+}(t), t \in \mathbb{N}\}$, is a random walk with drift equal to D_0, defined in (9). Thus, by Wald's identity it follows that

$$D_0 \; E_0[\tau] = E_0[-Z_{e+}(\tau)]. \tag{16}$$

By the information theoretic inequality in [19, Chapter 3.2] we have

$$\begin{aligned} E_0[-Z_{e+}(\tau)] &= E_0\left[\log \frac{dP_0}{dP_{e+}}(\mathcal{F}_\tau)\right] \\ &\geq h(P_0(d \neq \emptyset), P_{e+}(d = \emptyset)). \end{aligned}$$

By the definition of $\Delta(\alpha, \beta, \gamma)$ we have $P_0(d \neq \emptyset) \leq \alpha$ and $P_{e+}(d = \emptyset) \leq \beta$. Since the function $h(x, y)$ is decreasing on the set $\{(x, y) : x + y \leq 1\}$, and by assumption $\alpha + \beta \leq 1$, we conclude that

$$E_0[-Z_{e+}(\tau)] \geq h(\alpha, \beta),$$

which along with (16) proves the first lower bound.

For the proof of the second lower bound, without loss of generality we focus on P_{e+}. Under P_{e+}, the process $\{Z_{e+}(t), t \in \mathbb{N}\}$ is a random walk with drift equal to D_1, and in a similar way as before we obtain

$$\begin{aligned} E_{e+}[Z_{e+}(\tau)] &= E_{e+}\left[\log \frac{dP_{e+}}{dP_0}(\mathcal{F}_\tau)\right] \\ &\geq h(P_{e+}(d = \emptyset), P_0(d \neq \emptyset)) \geq h(\beta, \alpha). \end{aligned}$$

Therefore, these two inequalities yield

$$E_{e+}[\tau] \geq \frac{h(\beta, \alpha)}{D_1}. \tag{17}$$

For any pair $e' \neq e$, consider the log-likelihood process

$$\log \frac{dP_{e+}}{dP_{e'+}}(\mathcal{F}_t) = Z_{e+}(t) - Z_{e'+}(t), \quad t \in \mathbb{N},$$

which is a random walk with drift $D_0 + D_1$ under P_{e+}. Note that

$$\{d_e = 0\} = \{d = \emptyset\} \cup \{d \neq \emptyset, d \neq \{e\}\}, \tag{18}$$

and consequently $P_{e+}(d_e = 0) \leq \beta + \gamma$. Therefore, as before we obtain

$$\begin{aligned} E_{e+}\left[\log \frac{dP_{e+}}{dP_{e'+}}(\mathcal{F}_\tau)\right] &\geq h(P_{e+}(d_e = 0), P_{e'+}(d_e = 1)) \\ &\geq h(\beta + \gamma, \gamma). \end{aligned}$$

Using $d_{e'}$ instead of d_e, we can similarly show that

$$\begin{aligned} E_{e+}\left[\log \frac{dP_{e+}}{dP_{e'+}}(\mathcal{F}_\tau)\right] &\geq h(P_{e+}(d_{e'} = 1), P_{e'+}(d_{e'} = 0)) \\ &\geq h(\gamma, \beta + \gamma), \end{aligned}$$

which yields

$$E_{e+}[\tau] \geq \frac{h(\beta + \gamma, \gamma) \vee h(\gamma, \beta + \gamma)}{D_0 + D_1}. \tag{19}$$

Combining (17) and (19) completes the proof of the second lower bound. ■

The following lemma establishes an asymptotic upper bound on the expected sample size of the proposed testing procedure.

Lemma 4.2: Let $e \in \mathcal{E}$. As $A, B, C \to \infty$ we have

$$\begin{aligned} E_0[\tau_*] &\leq \frac{\log A}{D_0} \; (1 + o(1)) \\ E_{e-}[\tau_*], \; E_{e+}[\tau_*] &\leq \left(\frac{\log B}{D_1} \bigvee \frac{\log C}{D_0 + D_1}\right) \; (1 + o(1)). \end{aligned}$$

Proof: We will prove the second asymptotic lower bound, since the proof for the first one is similar. Moreover, without loss of generality we focus on the proof under P_{e+}. In view of (7)-(8), $\tau^* \leq \tau_2 \leq \tau_2' \leq \tau_2''$, where

$$\begin{aligned} \tau_2' = \inf\{n \geq 1 : \; & Z_e(n) \geq b \text{ and} \\ & Z_e(n) - Z_{e'}(n) \geq c \quad \forall \, e' \neq e\}, \\ \tau_2'' = \inf\{n \geq 1 : \; & Z_{e+}(n) \geq b + \log(2) \text{ and} \\ & Z_{e+}(n) - Z_{e'+}(n) - \xi_{e'}(n) \geq c \quad \forall \, e' \neq e\}, \end{aligned}$$

and $b := \log B$, $c := \log C$. Therefore, it suffices to establish the asymptotic upper bound for τ_2''. To this end, it suffices to prove that for any given $\epsilon > 0$ we have

$$\begin{aligned} &\sum_{n=1}^{\infty} P_{e+}(Z_{e+}(n) < n(D_1 - \epsilon)) < \infty \\ &\sum_{n=1}^{\infty} P_{e+}(Z_{e+}(n) - Z_{e'+}(n) < n(D_1 + D_0 - \epsilon/2)) < \infty \\ &\sum_{n=1}^{\infty} P_{e+}(-\xi_{e'}(n) < -n\epsilon/2) < \infty. \end{aligned}$$

for every $e' \neq e$. Fix $\epsilon > 0$ and $e' \neq e$. The first two series clearly converge (see, e.g., [20, Theorem 1]). The same is true for the third one since for sufficiently large n we have

$$\begin{aligned} &P_{e+}\left(\frac{1}{n}\xi_{e'}(n) > \frac{\epsilon}{2}\right) \\ &= P_{e+}\left(e^{-\frac{1}{2}\sum_{t=1}^{n} \mathbf{X}_{e'}(t)^T(\Sigma_-^{-1} - \Sigma_+^{-1})\mathbf{X}_{e'}(t)} > e^{\frac{n\epsilon}{2}} - 1\right) \\ &\leq P_{e+}\left(e^{-\frac{1}{2}\sum_{t=1}^{n} \mathbf{X}_{e'}(t)^T(\Sigma_-^{-1} - \Sigma_+^{-1})\mathbf{X}_{e'}(t)} > e^{\frac{n\epsilon}{4}}\right) \\ &\leq P_{e+}\left(\sum_{t=1}^{n} \mathbf{X}_{e'}(t)^T(\Sigma_-^{-1} - \Sigma_+^{-1})\mathbf{X}_{e'}(t) > \frac{n\epsilon}{2}\right) \\ &= \exp\left\{-n\frac{M^2\epsilon^2}{8 + 2M\epsilon}\right\}, \quad M \equiv \frac{1 - \rho_*^2}{8\rho_*} \end{aligned}$$

where the first inequality holds because $e^x - 1 > e^{x/2}, x \geq 1$, and the last one follows from the Hanson-Wright concentration inequality [21] for the distribution of a quadratic form of independent sub-Gaussian random variables. ■

Combining the two previous lemmas we can state and prove the main result of this paper.

Theorem 4.1: Suppose the thresholds in (τ_*, d_*) are selected according to (14). Then, for every $e \in \mathcal{E}$, as $\alpha, \beta, \gamma \to 0$ we have

$$E_0[\tau_*] \sim \inf_{(\tau,d)\in\Delta(\alpha,\beta,\gamma)} E_0[\tau] \sim \frac{|\log \beta|}{D_0}$$
$$E_{e+}[\tau_*] \sim \inf_{(\tau,d)\in\Delta(\alpha,\beta,\gamma)} E_{e+}[\tau] \sim \frac{|\log \alpha|}{D_1} \bigvee \frac{|\log \gamma|}{D_0 + D_1}$$
$$E_{e-}[\tau_*] \sim \inf_{(\tau,d)\in\Delta(\alpha,\beta,\gamma)} E_{e-}[\tau] \sim \frac{|\log \alpha|}{D_1} \bigvee \frac{|\log \gamma|}{D_0 + D_1}.$$

Proof: We prove the result for P_{e+} and some arbitrary $e \in \mathcal{E}$, since the other cases can be shown similarly. If the thresholds are selected according to (14), then from Lemma 4.2 it follows that

$$E_{e+}[\tau_*] \leq \left(\frac{|\log \alpha|}{D_1} \bigvee \frac{|\log \gamma|}{D_0 + D_1}\right)(1 + o(1)).$$

Moreover, in view of the definition of the function h in (15). it is clear that as $x, y \to 0$

$$h(x, y) \sim |\log y| \quad \text{and} \quad h(x, y) \vee h(y, x) \sim |\log(x \wedge y)|.$$

Thus, by Lemma 4.1 it follows that as $\alpha, \beta, \gamma \to 0$

$$\inf_{(\tau,d)\in\Delta(\alpha,\beta,\gamma)} E_{e+}[\tau] \geq \left(\frac{|\log \alpha|}{D_1} \bigvee \frac{|\log \gamma|}{D_0 + D_1}\right)(1 + o(1)),$$

which completes the proof. ■

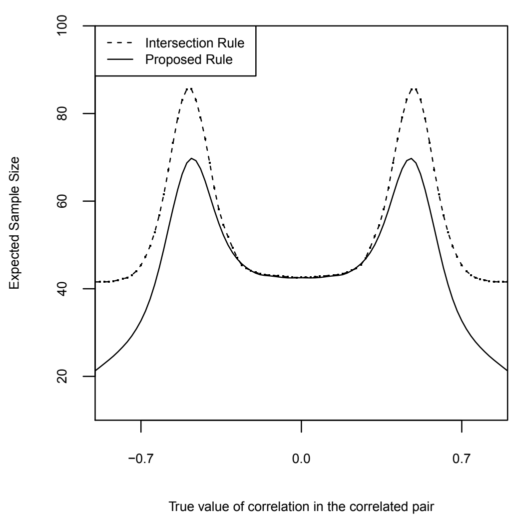

Figure 1. Expected Sample Size under different values of the correlation coefficient, ρ, in one pair, while all other pairs are uncorrelated

V. Comparison

In this section we present a simulation study in which we compare the proposed rule with a modification of the *Intersection Rule*, (τ_{int}, d_{int}), proposed in [17], according to which we stop as soon as either all log-likelihood ratio statistics are below $1/A$, or one of them is above B and all others below $1/A$, i.e.,

$$\tau_{int} := \inf\{n \geq 1 : 0 \leq p(n) \leq 1 \text{ and } \Lambda_e(n) \notin (1/A, B) \text{ for all } e \in \mathcal{E}\},$$
$$d_{int} := \begin{cases} \emptyset & \text{if } p(\tau_{int}) = 0 \\ i_1(\tau_{int}) & \text{otherwise.} \end{cases}$$

where $p(n)$ represents the number of mixture likelihood ratio statistics at time n that are greater than 1. It can be shown, similarly to Theorem 3.1, that $(\tau_{int}, d_{int}) \in \Delta(\alpha, \beta, \gamma)$ when the thresholds are selected as

$$A = \frac{1}{\beta} \quad \text{and} \quad B = \max\left\{\frac{K}{\alpha}, \frac{K-1}{\gamma}\right\}. \tag{20}$$

For our simulation study we set $p = 10, \rho_* = 0.7, \alpha = \beta = 10^{-2}$, $\gamma = 10^{-3}$. We select the thresholds for the proposed rule according to (14) and for the intersection rule according to (20). We estimate the expected sample size of both procedures when all pairs are uncorrelated apart from one that has correlation ρ. We consider different values for ρ in the interval $(-0.9, 0.9)$, not only the cases $\rho = 0$ and $\rho = \rho_*, -\rho_*$ that we considered in our theoretical results.

From Figure 1 we observe that when the absolute value of ρ is larger than ρ^* (resp. close to 0), the expected sample size of the proposed rule (τ_*, d_*) is much smaller than (resp. essentially the same as) that of (τ_{int}, d_{int}). Moreover, we observe that the worst-case scenario for both procedures occurs when $|\rho|$ is equal to some value between 0 and ρ^*. However, it is interesting to see that, even in this case, (τ_*, d_*) has a visibly better expected sample size than (τ_{int}, d_{int}).

VI. Acknowledgements

This work was supported in part by National Science Foundation through grant NSF ATD 1737962 and NSF AMPS 1736454.

References

[1] M. Schipper and E. Meelis, "Making sequential analysis of environmental monitoring data feasible by simplifying the covariance matrix structure," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 8, no. 1, pp. 122–137, 2003. [Online]. Available: http://www.jstor.org/stable/1400616

[2] Chin-Jen Ku and T. L. Fine, "Testing for stochastic independence: application to blind source separation," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1815–1826, May 2005.

[3] V. B. Balakirsky, A. R. Ghazaryan, and A. J. H. Vinck, "Testing the independence of two non-stationary random processes with applications to biometric authentication," in *2007 IEEE International Symposium on Information Theory*, June 2007, pp. 2671–2675.

[4] A. Anandkumar, L. Tong, and A. Swami, "Detection of gauss–markov random fields with nearest-neighbor dependency," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 816–827, Feb 2009.

[5] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *52nd IEEE Conference on Decision and Control*, Dec 2013, pp. 1659–1664.
[6] V. Solo and A. Pasha, "Testing for independence between a point process and an analog signal," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 3762–3765.
[7] S. C. Choi, "Sequential test for correlation coefficients," *Journal of the American Statistical Association*, vol. 66, no. 335, pp. 575–576, 1971. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1971.10482308
[8] C. J. Kowalski, "The oc and asn functions of some sprt's for the correlation coefficient," *Technometrics*, vol. 13, no. 4, pp. 833–841, 1971. [Online]. Available: http://www.jstor.org/stable/1266958
[9] M. Pradhan and Y. S. Sathe, "An unbiased estimator and a sequential test for the correlation coefficient," *Journal of the American Statistical Association*, vol. 70, no. 349, pp. 160–161, 1975. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1975.10480278
[10] G. Wolde-Tsadik, "A generalization of an sprt for the correlation coefficient," *Journal of the American Statistical Association*, vol. 71, no. 355, pp. 709–710, 1976. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10481552
[11] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945. [Online]. Available: http://www.jstor.org/stable/2235829
[12] M. Woodroofe, "Repeated likelihood ratio tests," *Biometrika*, vol. 66, no. 3, pp. 453–463, 12 1979. [Online]. Available: https://doi.org/10.1093/biomet/66.3.453
[13] J. Bartroff and T. L. Lai, "Multistage tests of multiple hypotheses," *Communications in Statistics - Theory and Methods*, vol. 39, no. 8-9, pp. 1597–1607, 2010. [Online]. Available: https://doi.org/10.1080/03610920802592852
[14] J. Bartroff and J. Song, "Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates," 2013.
[15] ——, "Sequential tests of multiple hypotheses controlling type i and ii familywise error rates," *Journal of Statistical Planning and Inference*, vol. 153, 04 2013.
[16] Y. Song and G. Fellouris, "Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals," *Electron. J. Statist.*, vol. 11, no. 1, pp. 338–363, 2017. [Online]. Available: https://doi.org/10.1214/17-EJS1223
[17] S. K. De and M. Baron, "Sequential bonferroni methods for multiple hypothesis testing with strong control of family-wise error rates i and ii," *Sequential Analysis*, vol. 31, no. 2, pp. 238–262, 2012. [Online]. Available: https://doi.org/10.1080/07474946.2012.665730
[18] J. Heydari and A. Tajer, "Quickest search for local structures in random graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 526–538, Sep. 2017.
[19] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, 08 2014.
[20] M. L. Katz, "The probability in the tail of a distribution," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 312–318, 03 1963. [Online]. Available: https://doi.org/10.1214/aoms/1177704268
[21] D. L. Hanson and F. T. Wright, "A bound on tail probabilities for quadratic forms in independent random variables," *Ann. Math. Statist.*, vol. 42, no. 3, pp. 1079–1083, 06 1971. [Online]. Available: https://doi.org/10.1214/aoms/1177693335